

# Sufficient conditions of endemic threshold on metapopulation networks

Taro Takaguchi<sup>1,2,†</sup> and Renaud Lambiotte<sup>3,‡</sup>

<sup>1</sup> National Institute of Informatics,  
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

<sup>2</sup> JST, ERATO, Kawarabayashi Large Graph Project,  
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

<sup>3</sup> Department of Mathematics and naXys, University of Namur,  
5000 Namur, Belgium

<sup>†</sup>t\_takaguchi@nii.ac.jp  
<sup>‡</sup>renaud.lambiotte@unamur.be

June 11, 2015

©2015, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives  
4.0 International <http://creativecommons.org/licenses/by-nc-nd/4.0/>

## Abstract

In this paper, we focus on susceptible-infected-susceptible dynamics on metapopulation networks, where nodes represent subpopulations, and where agents diffuse and interact. Recent studies suggest that heterogeneous network structure between elements plays an important role in determining the threshold of infection rate at the onset of epidemics, a fundamental quantity governing the epidemic dynamics. We consider the general case in which the infection rate at each node depends on its population size, as shown in recent empirical observations. We first prove that a sufficient condition for the endemic threshold (*i.e.*, its upper bound), previously derived based on a mean-field approximation of network structure, also holds true for arbitrary networks. We also derive an improved condition showing that networks with the rich-club property (*i.e.*, high connectivity between nodes with a large number of links) are more prone to disease spreading. The dependency of infection rate on population size introduces a considerable difference between this upper bound and estimates based on mean-field approximations, even when degree-degree correlations are considered. We verify the theoretical results with numerical simulations.

Keywords: Infectious diseases; Complex networks; Epidemiology

# 1 Introduction

Spreading of infectious diseases is one of the major threats to human society and understanding their dynamics is crucial to control them. When modeling the dynamics of infectious diseases in large-scale populations, it is often useful to adopt a metapopulation approach [1–3]. The detailed, most of the time unknown, network of contacts between individuals is then replaced by a coarse-grained network, where nodes represent regions in which individuals interact, *e.g.*, cities, and where links correspond to flows of individuals between the nodes. The computational and mathematical analysis of metapopulation models aims at identifying how network topology and model parameters affect the global epidemics. In a majority of realistic scenarios, nodes of metapopulation networks tend to exhibit a strong heterogeneity in their size and their degree. For instance, in the case of network of cities, it is well known that the size of cities is distributed according to a power-law [4, 5], and that their total flow is well approximated by their size [6]. Recent research has mainly focused on the impact of the degree distribution on epidemic spreading [3, 7–11], but the effect of their size distribution remains, to our knowledge, poorly known.

Directly related to this problem, the dependency of the contact rate between individuals on the size of population is a central theme in mathematical epidemiology [12]. To date, most studies have used terms that are either population-dependent, where the contact rate varies linearly with population size, or population-independent, where the rate is constant. In the literature, directly transmitted diseases such as measles, and foot and mouth disease, tend to be described by population-dependent models [13, 14], whereas sexually transmitted diseases such as HIV are usually considered population-independent [15]. More importantly, studies have shown that models parameterized with different transmission terms can predict very different quantitative and qualitative infection dynamics [16]. These observations suggest an intermediate scenario, where the scaling of contact rate on population size takes the general form  $N^\gamma$  with  $0 \leq \gamma \leq 1$ . In this direction, empirical measures on rodents have shown that

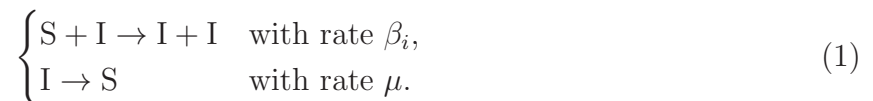
the standard modeling parameters,  $\gamma = 0$  and  $\gamma = 1$  can be significantly rejected, and that the observed scaling exponent lies between these two extremes [12]. In a different context, in the urban sciences, where the size of cities is known to play a fundamental role for collective social behaviors [17], recent research based on mobile phone data has also shown that the total number of contacts grows super-linearly with city population size [18]. Taken together, these findings motivate the study of metapopulation models with general contact rates, and a better understanding of the impact of city size distribution on epidemic spreading.

## 2 Model

### 2.1 Definition of the reaction–diffusion process

We consider a metapopulation network (we simply call it the network hereafter), composed of  $N$  nodes with  $n_p$  particles on it. Particles move between nodes through the links and nodes serve as substrate of interaction between particles.

In this paper, we assume that particles interact with each other according to so-called susceptible-infected-susceptible (SIS) dynamics [1], in line with the preceding works [3, 7–11, 19–21]. In the SIS dynamics, a particle takes either of two states, *i.e.*, susceptible (S) and infected (I), and the infection and recovery processes are described as follows:



Here we assume that the infection rate  $\beta_i$  is different for every node  $i$  ( $1 \leq i \leq N$ ). Aside from the reactions, particles perform diffusion between nodes with diffusion coefficient  $D_S$  and  $D_I$  for susceptible and infected particles, respectively. In general, SIS dynamics has an absorbing state corresponding to the disease-free equilibrium where there is no infected particle. The SIS dynamics exhibits a phase transition at a threshold value  $\beta_c$  such that in the case of  $\beta > \beta_c$  the endemic equilibrium emerges and there is a finite fraction of infected particles at the stationary state [22]. The estimation of  $\beta_c$  for arbitrary metapopulation networks is the main concern of the present study.

We assume that interaction between particles within a node and diffusion of particles between nodes occur simultaneously [7], instead of the two-step process of reaction and diffusion investigated in [3]. In the continuous time limit, the average behavior of the dynamics can be described by the deterministic reaction-diffusion equation on an arbitrary network [19] as

$$\partial_t \rho_{S,i} = \mu \rho_{I,i} - \frac{\beta_i}{\rho_i} \rho_{S,i} \rho_{I,i} - D_S \sum_j L_{ij} \rho_{S,j}, \quad (2)$$

$$\partial_t \rho_{I,i} = -\mu \rho_{I,i} + \frac{\beta_i}{\rho_i} \rho_{S,i} \rho_{I,i} - D_I \sum_j L_{ij} \rho_{I,j}, \quad (3)$$

where  $\rho_{S,i}$  and  $\rho_{I,i}$  represent the number of susceptible and infected particles at node  $i$ , respectively. The total number of particles at node  $i$ , denoted by  $\rho_i$ , is equal to  $\rho_{S,i} + \rho_{I,i}$ . Matrix  $L$  denotes the random-walk Laplacian whose elements are given by

$$L_{ij} = \begin{cases} 1 & (i = j), \\ -A_{ij}/k_j & (i \neq j), \end{cases} \quad (4)$$

where  $A_{ij}$  is the element of adjacency matrix  $A$  and equal to unity if nodes  $i$  and  $j$  are adjacent and equal to zero otherwise. Degree  $k_i$  of node  $i$  represents the number of links connected to the node. In this paper, we assume that the network is connected (*i.e.*, consists of a single connected component) via undirected and unweighted links and is simple. In other words,  $A$  is a symmetric matrix, with diagonal elements  $A_{ii} = 0$  ( $1 \leq i \leq N$ ). It should be noted that the total number of particles is conserved throughout the process. The average number of particles on a node is denoted by  $\rho \equiv n_p/N$ .

The infection rate at node  $i$  is set to depend on  $\rho_i$  as a power-law scaling given by  $\beta_i \equiv \beta \rho_i^\gamma$  with  $\gamma \geq 0$ . This scaling, motivated by recent empirical findings [12, 18] described in the introduction, was also considered in Refs. [11, 16, 23]. A special case  $\gamma = 0$  corresponds to the density-dependent SIS dynamics [3] and  $\gamma = 1$  the so-called mass interaction which was considered in Refs. [3, 19]. Instead of considering a binary setting where  $\gamma \sim 1$  for nodes with a small population and  $\gamma \sim 0$  for nodes with a large population [21], we assume the same  $\gamma$  for all nodes such that infection rate continuously increases with population size.

## 2.2 Heterogeneous mean-field approximation

In this section, we briefly describe the known results about the dynamics under the heterogeneous mean-field (HMF) approximation [7, 10]. Under the HMF approximation, the nodes with the same degree are assumed to behave in the same way on average. To be more precise, equations (2) and (3) are reduced to

$$\partial_t \rho_{S,k} = \mu \rho_{I,k} - \beta \rho_k^{\gamma-1} \rho_{S,k} \rho_{I,k} - D_S \rho_{S,k} + D_S k \sum_{k'} P(k'|k) \frac{1}{k'} \rho_{S,k'}, \quad (5)$$

$$\partial_t \rho_{I,k} = -\mu \rho_{I,k} + \beta \rho_k^{\gamma-1} \rho_{S,k} \rho_{I,k} - D_I \rho_{I,k} + D_I k \sum_{k'} P(k'|k) \frac{1}{k'} \rho_{I,k'}. \quad (6)$$

In uncorrelated networks,  $P(k'|k) = k'P(k')/\langle k \rangle$  holds true (where  $\langle k \rangle$  is the average degree of the network) and the equations are simplified as

$$\partial_t \rho_{S,k} = \mu \rho_{I,k} - \beta \rho_k^{\gamma-1} \rho_{S,k} \rho_{I,k} - D_S \left( \rho_{S,k} - \frac{k}{\langle k \rangle} \rho_S \right), \quad (7)$$

$$\partial_t \rho_{I,k} = -\mu \rho_{I,k} + \beta \rho_k^{\gamma-1} \rho_{S,k} \rho_{I,k} - D_I \left( \rho_{I,k} - \frac{k}{\langle k \rangle} \rho_I \right), \quad (8)$$

where  $\rho_S \equiv \sum_k P(k) \rho_{S,k}$  and  $\rho_I \equiv \sum_k P(k) \rho_{I,k}$ .

We are interested in the condition of parameters for realizing the endemic equilibrium in the long-term limit (*i.e.*,  $t \rightarrow \infty$  and  $\partial_t \rho_{S,k}, \partial_t \rho_{I,k} \rightarrow 0$ ). In this limit, if  $D_S, D_I > 0$ ,  $\rho_k$  converges to the stationary distribution  $\rho_k^*$  regardless of the other parameters. Equations (7) and (8) have a fixed point  $(\rho_{S,k}^*, \rho_{I,k}^*) = (k\rho/\langle k \rangle, 0)$  which corresponds to the disease-free equilibrium. The endemic equilibrium arises when the disease-free equilibrium is unstable. The linear stability of the disease-free equilibrium is determined by the Jacobian matrix at the fixed point:

$$J_{\text{HMF}} = \begin{pmatrix} J_{\text{HMF}}^{(1)} & J_{\text{HMF}}^{(3)} \\ O & J_{\text{HMF}}^{(2)} \end{pmatrix}, \quad (9)$$

where  $J_{\text{HMF}}^{(i)}$  ( $i = 1, 2, 3$ ) are  $k_{\max} \times k_{\max}$  matrices and  $O$  is the  $k_{\max} \times k_{\max}$  zero matrix, where

$k_{\max}$  is the maximum  $k_i$  in the network. The elements of  $J_{\text{HMF}}^{(1)}$  and  $J_{\text{HMF}}^{(2)}$  are given by

$$\left(J_{\text{HMF}}^{(1)}\right)_{kk'} = -D_S \left( \delta_{k,k'} + \frac{k}{\langle k \rangle} P(k') \right), \quad (10)$$

$$\left(J_{\text{HMF}}^{(2)}\right)_{kk'} = \left( -\mu + \beta \left( \frac{k}{\langle k \rangle} \rho \right)^\gamma - D_I \right) \delta_{k,k'} + D_I \frac{k}{\langle k \rangle} P(k'), \quad (11)$$

where  $\delta_{kk'}$  denotes the Kronecker delta.

The disease-free equilibrium becomes unstable when the real part of the largest eigenvalue of  $J_{\text{HMF}}$  is positive. As we see in Eq. (9), the eigenvalues of  $J_{\text{HMF}}$  are given by the union of the eigenvalues of  $J_{\text{HMF}}^{(1)}$  and  $J_{\text{HMF}}^{(2)}$ . From Eq. (10),  $J_{\text{HMF}}^{(1)}$  consists of the sum of identity matrix and a rank-1 matrix whose eigenvalues are zero (with degeneracy factor  $k_{\max} - 1$ ) and unity. Therefore, the eigenvalues of  $J_{\text{HMF}}^{(1)}$  are equal to  $-D_S$  and  $-2D_S$  and negative regardless of parameter values. Then, the largest eigenvalue of  $J_{\text{HMF}}$  is given by the largest eigenvalue of  $J_{\text{HMF}}^{(2)}$ . The largest eigenvalue of  $J_{\text{HMF}}^{(2)}$  satisfies

$$\max_{1 \leq k \leq k_{\max}} \left( -\mu + \beta \left( \frac{k}{\langle k \rangle} \rho \right)^\gamma - D_I \right) \leq \lambda_{\max}(J_{\text{HMF}}^{(2)}) \leq \max_{1 \leq k \leq k_{\max}} \left( -\mu + \beta \left( \frac{k}{\langle k \rangle} \rho \right)^\gamma \right). \quad (12)$$

Therefore, the sufficient condition of  $\beta$  for positive  $\lambda_{\max}(J_{\text{HMF}}^{(2)})$  is given by

$$\beta > (\mu + D_I) \left( \frac{k_{\max}}{\langle k \rangle} \rho \right)^{-\gamma}. \quad (13)$$

and an upper bound of the endemic threshold  $\beta_c$  is derived as

$$\bar{\beta}_{\text{HMF}} \equiv (\mu + D_I) \left( \frac{k_{\max}}{\langle k \rangle} \rho \right)^{-\gamma} \quad (14)$$

This upper bound of endemic threshold is trivially determined by the most susceptible node, that is the most populated one where the infection rate is maximum in our setting. Although we only reviewed the analysis for uncorrelated networks, the same upper bound is also proven for correlated random networks, *i.e.*,  $P(k'|k) \neq k'P(k')/\langle k \rangle$  (see Theorem 4.3 in Ref. [11]).

In closing this section, it is worth to mention the relationship with the endemic threshold derived in the previous studies. When we set  $\gamma = 1$  and  $D_I = 1$  for the two-step reaction-diffusion modeling [3] and  $D_I \rightarrow \infty$  for the simultaneous reaction-diffusion modeling [19] (the

same as one we adopt in this paper), the endemic threshold  $\beta_c$  is exactly derived as

$$\beta_c = \frac{\langle k \rangle^2 \mu}{\langle k^2 \rangle \rho}. \quad (15)$$

This functional form of  $\beta_c$  implies that the heterogeneity in  $P(k)$  plays a crucial role in determining the critical point, as one observes in the percolation processes [24, 25] and epidemic processes on contact networks (*i.e.*, networks between individuals) [26, 27]. The two studies reporting this  $\beta_c$  [3, 19] have an assumption in common on deriving Eq. (15); the diffusion of infected particles are very fast and thus  $\rho_{I,i} = k_i / \langle k \rangle \rho_I$  holds true. Under this assumption, the endemic threshold  $\beta_c$  for any value of  $\gamma$  is given by Eq. (15). If this assumption does not hold true, the exact  $\beta_c$  is no longer tractable and its upper bound is obtained as described above.

## 3 Results

### 3.1 Endemic threshold for arbitrary networks

#### 3.1.1 Proof of $\bar{\beta}_{\text{HMF}}$ for general case

We first prove that the upper bound  $\bar{\beta}_{\text{HMF}}$  (given by Eq. (14)) also holds true for arbitrary networks (see Eq. (26)). The following results are essentially in parallel with previous studies [7, 9–11, 19, 21]. The most important difference with these studies is that they assume a HMF approximation, except for Ref. [19], whereas we derive our results directly from the adjacency matrix of the network, thereby considering arbitrary networks free from this approximation.

As we saw in the derivation with the HMF approximation (see Sec. 2.2), we consider the stability of the disease-free equilibrium  $(\rho_{S,i}^*, \rho_{I,i}^*) = (k_i / \langle k \rangle \rho, 0)$  ( $1 \leq i \leq N$ ) of the original reaction-diffusion equations (given by Eqs. (2) and (3) with  $\beta_i \equiv \beta \rho_i^\gamma$ ). The linear stability of the disease-free equilibrium is determined by the Jacobian matrix at the equilibrium:

$$J = \begin{pmatrix} J^{(1)} & J^{(3)} \\ O & J^{(2)} \end{pmatrix} \quad (16)$$

All matrices  $J^{(1)}$ ,  $J^{(2)}$ ,  $J^{(3)}$ , and  $O$  are with size  $N \times N$ . Matrix  $O$  is zero matrix. The element



of  $J^{(1)}$  and  $J^{(2)}$  is given by

$$J_{ij}^{(1)} = -D_S L_{ij}, \quad (17)$$

$$J_{ij}^{(2)} = \left( \beta \left( \frac{k_i}{\langle k \rangle} \rho \right)^\gamma - \mu \right) \delta_{ij} - D_I L_{ij}. \quad (18)$$

Because the eigenvalues of  $L$  are in  $0 \leq \lambda(L) \leq 2$  [28], the eigenvalues of  $J^{(1)}$  are in  $-2D_S \leq \lambda(J^{(1)}) \leq 0$  and negative regardless of parameter values. Therefore, to derive the condition for the endemic equilibrium, we investigate the largest eigenvalue of  $J^{(2)}$ , denoted by  $\lambda_{\max}(J^{(2)})$ , in the following.

Before moving to general cases, we note two special cases in which we can exactly derive  $\beta_c$ . As the first case, when  $\gamma = 0$ , the element of  $J^{(2)}$  is equal to

$$J_{ij}^{(2)} = (\beta - \mu) \delta_{ij} - D_I L_{ij}. \quad (19)$$

Therefore,  $\lambda_{\max}(J^{(2)})$  is given by  $(\beta - \mu)$  and it leads to  $\beta_c = \mu$ . As the second case, when the network is regular and  $k_i = k$  ( $1 \leq i \leq N$ ), the element of  $J^{(2)}$  is equal to

$$J_{ij}^{(2)} = (\beta \rho^\gamma - \mu) \delta_{ij} - D_I L_{ij}. \quad (20)$$

Therefore,  $\lambda_{\max}(J^{(2)})$  is given by  $(\beta \rho^\gamma - \mu)$  and it leads to  $\beta_c = \mu \rho^{-\gamma}$ . It is worthy to note that the endemic threshold  $\beta_c$  is independent of diffusion rate  $D_I$  for these two cases.

Now we consider general cases with  $\gamma \neq 0$  on a non-regular network. For the convenience for calculation, we symmetrize  $J^{(2)}$  via the following similarity transformation [19]:

$$\hat{J}^{(2)} \equiv \text{diag} \left( \frac{1}{\sqrt{k_1}}, \frac{1}{\sqrt{k_2}}, \dots, \frac{1}{\sqrt{k_N}} \right) J^{(2)} \text{diag} \left( \sqrt{k_1}, \sqrt{k_2}, \dots, \sqrt{k_N} \right), \quad (21)$$

where  $\text{diag}(a_1, a_2, \dots, a_N)$  represents a diagonal matrix whose  $(i, i)$ -element is equal to  $a_i$ . Element of  $\hat{J}^{(2)}$  is given by

$$\hat{J}_{ij}^{(2)} = \left( \beta \left( \frac{k_i}{\langle k \rangle} \rho \right)^\gamma - \mu - D_I \right) \delta_{ij} + D_I \frac{A_{ij}}{\sqrt{k_i k_j}}. \quad (22)$$

When the largest eigenvalue of  $\hat{J}^{(2)}$ , denoted by  $\lambda_{\max}(\hat{J}^{(2)})$ , is positive, the disease-free equilibrium is unstable and the endemic equilibrium arises. As we did for the case with HMF

approximation (see Sec. 2.2), we estimate a lower bound of  $\lambda_{\max}(\hat{J}^{(2)})$  (that leads to an upper bound of  $\beta_c$ ). By using the Rayleigh quotient,  $\lambda_{\max}(\hat{J}^{(2)})$  is given by

$$\lambda_{\max}(\hat{J}^{(2)}) = \max_{|x|=1} \frac{x^\top \hat{J}^{(2)} x}{x^\top x}. \quad (23)$$

We show that the upper bound for HMF approximation can be easily recovered as follows. Let us set  $x = (0, 0, \dots, 1, \dots, 0)^\top$  in which only the  $i$ -th element of  $x$  is equal to unity and the rest equal to zero. With this  $x$ ,

$$x^\top \hat{J}^{(2)} x = \beta \left( \frac{k_i}{\langle k \rangle} \rho \right)^\gamma - \mu - D_I. \quad (24)$$

Therefore,

$$\lambda_{\max}(\hat{J}^{(2)}) \geq \max_{1 \leq i \leq N} \beta \left( \frac{k_i}{\langle k \rangle} \rho \right)^\gamma - \mu - D_I, \quad (25)$$

and we obtain the sufficient condition of  $\beta$  as

$$\beta > (\mu + D_I) \left( \frac{k_{\max}}{\langle k \rangle} \rho \right)^{-\gamma} = \bar{\beta}_{\text{HMF}}. \quad (26)$$

Notice that this sufficient condition for the endemic state is equivalent to that for HMF approximation given by Eq. (14) [7, 9–11].

### 3.1.2 Improved upper bound and the rich-club phenomenon

In this section, we show an improved upper bound of  $\beta_c$  and argue its relationship with the so-called rich-club phenomenon observed in real networks. Our idea of improvement stems from the observation that the upper bound given by Eq. (26) is achieved via approximation of the principal eigenvector of  $\hat{J}^{(2)}$  with a vector of zeros except for the element corresponding to a single node with  $k_{\max}$ . We relax the criterion about node degree and approximate the principal eigenvector with a normalized vector whose element increases with  $k_i$  of the corresponding nodes. Let us set  $x$  such that

$$x_i = \begin{cases} 0 & (k_i < k_c), \\ \sqrt{\frac{k_i}{K_c}} & (k_i \geq k_c), \end{cases} \quad (27)$$

where the free parameter  $k_c$  takes an integer value in  $[1, k_{\max}]$  and  $K_c \equiv \sum_{j: k_j \geq k_c} k_j$  is a normalization constant. With this choice of  $x$ ,

$$x^\top \hat{J}^{(2)} x = \sum_{i: k_i \geq k_c} \beta \left( \frac{k_i}{\langle k \rangle} \rho \right)^\gamma \frac{k_i}{K_c} - \mu - D_I \left( 1 - \frac{2E_c}{K_c} \right), \quad (28)$$

$$\geq \beta \left( \frac{k_c}{\langle k \rangle} \rho \right)^\gamma - \mu - D_I \left( 1 - \frac{2E_c}{K_c} \right), \quad (29)$$

where  $E_c$  is the total number of links between nodes with  $k_i \geq k_c$ . Therefore,

$$\lambda_{\max}(\hat{J}^{(2)}) = \max_{|x|=1} \frac{x^\top \hat{J}^{(2)} x}{x^\top x} \geq \max_{k_c} \left[ \beta \left( \frac{k_c}{\langle k \rangle} \rho \right)^\gamma + D_I \frac{2E_c}{K_c} \right] - (\mu + D_I). \quad (30)$$

A sufficient condition for the endemic equilibrium is that there exists  $k_c$  which satisfies

$$\beta \left( \frac{k_c}{\langle k \rangle} \rho \right)^\gamma - \left( 1 - \frac{2E_c}{K_c} \right) D_I - \mu > 0. \quad (31)$$

In terms of  $\beta$ , this is equivalent to

$$\beta > \bar{\beta} \equiv \min_{k_c} \left\{ \left[ \mu + \left( 1 - \frac{2E_c}{K_c} \right) D_I \right] \left( \frac{k_c}{\langle k \rangle} \rho \right)^{-\gamma} \right\}. \quad (32)$$

This sufficient condition always improves the previous one (given by Eq. (26)) in the following sense. Let us focus on the right hand side of Eq. (32) with  $k_c = k_{\max}$ . If there is no link between the nodes with  $k_i = k_{\max}$ , then  $E_c = 0$  and we recover Eq. (26). Otherwise, if  $E_c > 0$ , the r.h.s. of Eq. (32) is smaller than the r.h.s. of Eq. (26) and we obtain a smaller upper bound. We denote the right hand side of Eq. (32) by  $\bar{\beta}$ , as an improved upper bound of endemic threshold to be compared with  $\bar{\beta}_{\text{HMF}}$ .

The expression on the right hand side of Eq. (32) is closely related to so-called rich-club phenomenon of networks [29, 30], characterized with a high connectivity between nodes with large degree, and to the presence of a  $k$ -core [31], as we discuss further below. Formally, the rich-club phenomenon is defined by focusing on [29, 30]

$$\phi(k) \equiv \frac{2E_{>k}}{N_{>k}(N_{>k} - 1)}, \quad (33)$$

where  $N_{>k}$  is the number of nodes with degree larger than  $k$  and  $E_{>k}$  is the total number of links between these nodes. The value of  $\phi(k)$  for the original network is usually compared to  $\phi_{\text{rand}}(k)$  of a network generated by randomly rewiring links while preserving degree of each node. If  $\phi(k) > \phi_{\text{rand}}(k)$  for large values of  $k$ , the network is said to exhibit the rich-club phenomenon [30]. The factor  $2E_c/K_c$  which appears in Eq. (32) is similar to  $\phi(k)$  and we expect that  $\bar{\beta}$  would estimate  $\beta_c$  better especially when the network has the rich-club structure (see Sec. 3.2.4 for a real network example). It should be noted that our derivation of Eq. (32) results from the stability analysis of the dynamical process while  $\phi(k)$  is purely based on network structure. The sufficient condition  $\bar{\beta}$  depends on the rate of contacts inside high degree nodes, as previously, but also on the structure of the network, via the existence of links between these high-degree nodes. The extent to which  $\bar{\beta}$  improves  $\bar{\beta}_{\text{HMF}}$  depends on the underlying network. It is also important to stress that structure becomes more and more important as  $D_I$  is increased, and the disease rapidly mixes in the population.

It is worth noting the relationship between our upper bound  $\bar{\beta}$  given by Eq. (32) and another upper bound recently derived based on a mean-field approximation taking into account the correlation of degrees between adjacent nodes [11],

$$\beta > \bar{\beta}_{\text{MFC}} \equiv \min_{k_i} \left\{ [\mu + (1 - P(k_i|k_i)) D_I] \left( \frac{k_i}{\langle k \rangle} \rho \right)^{-\gamma} \right\}, \quad (34)$$

where the subscript MFC stands for mean-field approximation with correlated structure. If  $P(k_c|k_c) = 2E_c/K_c$  holds true for the  $k_c$  minimizing both Eqs. (32) and (34),  $\bar{\beta}_{\text{MFC}}$  is equivalent to  $\bar{\beta}$ . The accuracy of each upper bound is expected to vary depending on the structural correlations present in the network. It should be noted that the derivation of  $\bar{\beta}_{\text{MFC}}$  [11] assumes a mean-field approximation of the dynamical process, where nodes with the same  $k_i$  value behave in the same manner (as we saw in Eqs. (5) and (6) in Sec. 2.2). In contrast, our derivation of  $\bar{\beta}$  is directly derived from Eq.(3) and is free from such an approximation. It should also be noted that  $\bar{\beta}_{\text{MFC}}$  does not take into account, by definition, structural correlations beyond those of the degrees between adjacent nodes, and that  $\bar{\beta}_{\text{MFC}}$  is determined by point correlation

$P(k_i|k_i)$ . The upper bound  $\bar{\beta}$  is also affected by degree correlations, but it depends on the global structure of the network, as the rich-club phenomenon, because correlations are considered in a range of  $k_i$ . We will compare the accuracy of  $\bar{\beta}$  and  $\bar{\beta}_{\text{MFC}}$  in the next section for both synthetic and empirical networks.

### 3.2 Numerical results

In this section, we investigate the accuracy of the upper bounds of  $\beta_c$  (that we obtained and reviewed in Sec. 3.1) by comparing them with numerical simulations of the epidemic dynamics. For the sake of simplicity, we assume  $D_S = D_I$  and fix  $(\mu, \rho) = (1, 100)$ . We focus on three parameters  $(\beta, \gamma, D_I)$  and calculate  $f_I^* \equiv \sum_i \rho_{I,i}^*/n_p$ , which is the fraction of infected particles at the steady state, for different parameter values. For all numerical simulations, we set the initial condition to  $(\rho_{S,i}, \rho_{I,i}) = (k_i/\langle k \rangle \rho, 0)$  ( $1 \leq i \leq N$ ) except for the initially infected node  $j$  which is chosen at random and assigned with  $\rho_{I,j} = 4$  and  $\rho_j = k_j/\langle k \rangle \rho$ . Even above the endemic threshold, the SIS dynamics can hit the absorbing state, *i.e.*, the disease-free equilibrium with non-zero probability, due to stochastic fluctuations. To circumvent this issue and compute the stationary state, we adapt the so-called quasi-stationary simulation method [32–34] in which, every time the system reaches the absorbing state, the configuration of the nodes' states is replaced with one chosen randomly from the history. We also implement the simulation as it is done for migration processes [35], by not distinguishing particles and keeping only  $(\rho_{S,i}, \rho_{I,i})$  for each node. The time evolution is simulated as a set of transitions of the Markov chain in which a state corresponds to a configuration  $\{(\rho_{S,1}, \rho_{I,1}), \dots, (\rho_{S,N}, \rho_{I,N})\}$  [35]. For each combination of parameters' values, we calculate  $f_I^*$  as follows. We run the epidemic process from  $t = 0$  to  $t = T \equiv 1000/D_I$ , which is 1000 times longer than the characteristic time scale of particles' migration. In this period,  $\rho_i$  converges to its stationary distribution. Then we take average of  $f_I(t) \equiv \sum_i \rho_{I,i}(t)/n_p$  over  $T/2 \leq t \leq T$  and regard this average as the stationary value  $f_I^*$ .

This section consists of the following four parts. In Sec. 3.2.1, we consider the cases of a random regular graph with  $k = 4$  and a square lattice in order to confirm that regular

networks exhibit the same behavior, independently of their internal organization. In Sec. 3.2.2, we consider the case of the wheel graph in order to show that the sufficient condition holds true for a network whose structure differs from that of random graphs. In Sec. 3.2.3, we consider the case of random bimodal graphs with degree correlations, where  $k_i$  takes either of two values and where the degree correlation is given by  $P(k_i|k_j)$ , in order to systematically investigate the accuracy of the upper bounds in the presence of correlations. Finally, in Sec. 3.2.4, we consider the case of the US airports network, and show that the endemic threshold for a real-world rich-club network is closer to  $\bar{\beta}$  than to  $\bar{\beta}_{\text{HMF}}$  and  $\bar{\beta}_{\text{MFC}}$ .

### 3.2.1 Random regular graph and square lattice

We generate an instance of the random regular graph with degree  $k = 4$  with the use of configuration model [36, 37]. We also use the square lattice with periodic boundary condition in both directions, *i.e.*, the lattice on the surface of a torus. We set  $N = 1024$  for both networks. In Fig. 1, the fraction of infected particles at the stationary state, *i.e.*,  $f_I^*$ , is plotted with color maps as a function of focal parameters  $(\beta, \gamma, D_I)$ . Figures 1(a) and 1(b) show the results on the  $(\beta, \gamma)$ -plane. Although we observe deviation in the  $f_I^*$  values due to the fluctuation of stochastic simulations, the theoretical prediction of  $\beta_c = \mu\rho^{-\gamma}$  (drawn as the dotted line) fits well with the simulation results. Figures 1(c) and 1(d) show the results on the  $(\beta, D_I)$ -plane. The simulation results indicate that the onset of the endemic state does not depend on  $D_I$ , which is consistent with the theoretical prediction  $\beta_c = \mu\rho^{-\gamma}$ . In those figures, the regular random graph and square lattice are quantitatively indistinguishable, despite the fact that the former is appropriate for a mean-field approximation and the latter not.

### 3.2.2 Wheel graph

The wheel graph with  $N$  nodes consists of a cycle with  $N - 1$  nodes and a single hub connected to the rest of the nodes. Nodes on the cycle have degree 3 and the node at the center has degree  $N - 1$ . The average degree of the  $N$ -wheel graph is equal to  $\langle k \rangle = 4(N - 1)/N$ . For the wheel graph,  $\bar{\beta} = \bar{\beta}_{\text{HMF}} = \bar{\beta}_{\text{MFC}}$  holds true and are equal to  $(\mu + D_I) ((N - 1)/\langle k \rangle \rho)^{-\gamma}$  if

$N \gg 1$ . In Fig. 2(a),  $f_I^*$  is plotted on the  $(\beta, \gamma)$ -plane with  $D_I = 1$ . The upper bound is shown as the dashed line and is very accurate except for  $0 \leq \gamma \lesssim 0.2$ . In Fig. 2(b),  $f_I^*$  is plotted on the  $(\beta, D_I)$ -plane with  $\gamma = 0.5$ . The upper bound is shown as the dashed line again but is not tight when  $0.6 \lesssim D < 1$ . As a summary, the sufficient conditions given by Eqs. (26), (32), and (34) provide accurate description of the epidemic threshold even for a network where a mean-field assumption is not expected to hold.

### 3.2.3 Random bimodal graphs with degree correlation

In the previous sections, we have focused on random graphs with no correlation and on more regular graphs. However, real-world networks are neither purely random nor deterministic but often exhibit structural correlations. A typical example of structural correlation is the correlation of the degrees between adjacent node pairs, so-called degree assortativity [38, 39]. If two adjacent nodes tend to have similar degree values, the network is called assortative. If the nodes with a large degree tend to be connected to the nodes with a small degree, the network is instead called disassortative. Empirical observations suggest that similar types of networks are characterized by similar levels of assortativity [37]. The degree assortativity is quantified by the degree assortativity coefficient  $r$  which is defined as the Pearson correlation coefficient of the degrees of adjacent nodes [38, 39] and has values in the range  $-1 \leq r \leq 1$ . A positive (negative)  $r$  value represents an assortative (disassortative) network. Determining the influence of degree assortativity on the accuracy of the upper bounds  $\bar{\beta}_{\text{HMF}}$ ,  $\bar{\beta}$ , and  $\bar{\beta}_{\text{MFC}}$  is the main purpose of this section.

In order to tune the assortativity present in the networks, we generate instances of random bimodal graph with degree correlations as follows. In the random bimodal graph, a node has its degree equal to either of two values denoted by  $k_1$  and  $k_2$  ( $k_1 > k_2$ ), with probability  $a$  and  $1 - a$ , respectively. The degree assortativity coefficient is determined by the parameter  $P(k_1|k_2)$  which is the probability of a random link to have a node with degree  $k_1$  as an end under the condition that the other end has  $k_2$  [40], and equal to  $r = 1 - (ak_1 + (1 - a)k_2)P(k_1|k_2)/ak_1$

(see Appendix A for the derivation). We also calculate the expecting values of  $\bar{\beta}$  and  $\bar{\beta}_{\text{MFC}}$  as (see Appendix B for the derivations),

$$\bar{\beta} = \min \left[ \left( \mu + (1-r) \frac{(1-a)k_2}{\langle k \rangle} D_I \right) \left( \frac{k_1}{\langle k \rangle} \rho \right)^{-\gamma}, \mu \left( \frac{k_2}{\langle k \rangle} \rho \right)^{-\gamma} \right], \quad (35)$$

$$\bar{\beta}_{\text{MFC}} = \min \left[ \left( \mu + (1-r) \frac{(1-a)k_2}{\langle k \rangle} D_I \right) \left( \frac{k_1}{\langle k \rangle} \rho \right)^{-\gamma}, \left( \mu + (1-r) \frac{ak_1}{\langle k \rangle} D_I \right) \left( \frac{k_2}{\langle k \rangle} \rho \right)^{-\gamma} \right]. \quad (36)$$

Because  $(1-r)ak_1D_I/\langle k \rangle \geq 0$ ,  $\bar{\beta} \leq \bar{\beta}_{\text{MFC}}$  holds true regardless of parameter settings. In other words,  $\bar{\beta}$  is equivalent to or better than  $\bar{\beta}_{\text{MFC}}$  as an upper bound of  $\beta_c$  for the random bimodal graph.

In Fig. 3,  $f_1^*$  is plotted as a function of  $\beta$  for random bimodal graphs with disassortative, neutral, and assortative degree correlations. We consider two values of  $\gamma$ ,  $\gamma = 0.1$  and  $0.5$ . For  $\gamma = 0.1$ , that is when the dependency of infection rate on population size is weak,  $\bar{\beta}$  provides a more accurate estimation of the endemic threshold than  $\bar{\beta}_{\text{MFC}}$  for the disassortative and neutral cases (Figs. 3(a) and 3(b)). For the assortative case,  $\bar{\beta}$  is less accurate than, but comparable with  $\bar{\beta}_{\text{MFC}}$  (Fig. 3(c)). In all three cases,  $\bar{\beta}_{\text{HMF}}$  gives the poorest performance. When  $\gamma = 0.5$  and the dependency of infection rate on population size is strong,  $\bar{\beta}$  gives less accurate estimation than  $\bar{\beta}_{\text{MFC}}$  for all the three cases (Figs. 3(d), 3(e), and 3(b)) and the difference between the two upper bounds increases with  $r$ . The accuracy of the two upper bounds increases with  $r$ , while  $\bar{\beta}_{\text{HMF}}$  remains the same regardless of the value of  $r$ . The increase in accuracy of  $\bar{\beta}_{\text{MFC}}$  with  $r$  is expected because  $P(k_1|k_1)$  and  $P(k_2|k_2)$  become larger when  $r$  increases.

The theoretical results predict that  $\bar{\beta} < \bar{\beta}_{\text{MFC}}$  for  $\gamma = 0.1$  and  $r \in \{-0.6, 0\}$  and that  $\bar{\beta} = \bar{\beta}_{\text{MFC}}$  for  $(\gamma, r) = (0.1, 0.6)$  and  $\gamma = 0.5$  (see Fig. B1 in Appendix B). The small discrepancy between the theoretical predictions and the numerical results shown in Fig. 3 is probably due to the assumption behind the theory, as we assumed that the network is a typical instance of the random bimodal graph with a sufficiently large  $N$ . In contrast, in the numerical simulations, we calculate  $\bar{\beta}$  and  $\bar{\beta}_{\text{MFC}}$  for a single realization with  $N = 1000$ . Moreover, fluctuations in the node degree arise due to formation of self-loops and multiple links between a pair of nodes



when generating the graph. Because  $\bar{\beta}$  uses the exact value of degrees as the threshold, it is affected by such fluctuations. This observation suggests that  $\bar{\beta}$  theoretically provides a robust estimate in a broad variety of degree correlations, while its accuracy may be sensitive to the fluctuations due to randomness in network generation.

### 3.2.4 US airports network

In this section, we show simulation results on the US airports network [41, 42], a well-known example of rich-club network [30]. The system is a directed network in which nodes correspond to airports in the US and directed links between nodes represent directed flights between two airports in 2010 [41]. We discard the direction of all the links and regard it as an undirected network. We focus on its largest connected component with  $N = 1572$  nodes and  $E = 28235$  links. The degree assortativity coefficient  $r$  of the focal network is equal to  $-0.1134$ . The rich-club structure fosters the assortative mixing in the network, but it does not always lead to a positive  $r$  value of the whole network because the rich-club phenomenon pays attention only to the nodes with large  $k$  values. In this sense, the relationship between the degree correlation and rich-club phenomenon is not trivial. With numerical simulations, we examine to what extent  $\bar{\beta}$  improves  $\bar{\beta}_{\text{HMF}}$  and  $\bar{\beta}_{\text{MFC}}$  in this empirical network relevant to real-world epidemics.

In Fig. 4(a), we first evaluate the dependence of the r.h.s. of (32) and (34) on  $k$  for  $\gamma = 0.1$ . The minimum value of each curve corresponds to  $\bar{\beta}$  and  $\bar{\beta}_{\text{MFC}}$ , respectively. Because  $\bar{\beta}_{\text{HMF}}$  is independent of  $k$ , it is plotted as a horizontal line. While the r.h.s. of Eq. (32) follows a U-shape curve and takes its minimum at  $k = 17$ , the r.h.s. of Eq. (34) monotonically decreases with  $k$  and takes its minimum at  $k = k_{\text{max}} = 314$ , and its minimum value is equal to  $\bar{\beta}_{\text{HMF}}$ . Therefore, one finds  $\bar{\beta}_{\text{MFC}} = \bar{\beta}_{\text{HMF}}$  for the US airports network when  $\gamma = 0.1$ .

In Fig. 4(b),  $f_{\text{I}}^*$  is plotted as a function of infection rate  $\beta$  with  $\gamma = 0.1$ . The results of numerical simulations indicate  $\beta_{\text{c}} \sim 0.56$ . The vertical arrows point the three upper bounds  $\bar{\beta}$ ,  $\bar{\beta}_{\text{HMF}}$ , and  $\bar{\beta}_{\text{MFC}}$ , with  $\bar{\beta}_{\text{HMF}}$  and  $\bar{\beta}_{\text{MFC}}$  overlapping, as we saw in Fig. 4(a). It is clear that  $\bar{\beta}_{\text{HMF}}$  (equal to 0.967) is far from the value  $\beta_{\text{c}}$  measured in the simulations, whereas  $\bar{\beta}$  (equal to

0.729) drastically improves the estimation. These results support our main claim that rich-club networks have a smaller upper bound for  $\beta_c$  than non rich-club networks with the same degree sequence.

The difference between the upper bounds is then studied for values of  $\gamma$  in the range  $(0, 1]$ , as an exact expression of  $\beta_c$  can be derived for  $\gamma = 0$ , and that the three upper bounds are invalid for this value (see Sec. 3.1.1). In Fig. 4(c), we plot the values of  $\beta_c$  estimated from the simulations and the three upper bounds are plotted as a function of  $\gamma$ . We observe that  $\bar{\beta}_{\text{MFC}}$  is very close to  $\bar{\beta}_{\text{HMF}}$  over the whole the range of  $\gamma$  for this network. For  $0 < \gamma \lesssim 0.2$ ,  $\bar{\beta}$  is more accurate than  $\bar{\beta}_{\text{HMF}}$  and  $\bar{\beta}_{\text{MFC}}$  and closer to the simulation results, while the three upper bounds are equivalent for large values of  $\gamma$  (see the inset of Fig. 4(c)). Interestingly, real-world data tend to present small values of  $\gamma$  [12, 18], in the regime where  $\bar{\beta}$  improves most the accuracy of the estimate.

## 4 Discussion

In the present study, we have investigated the SIS epidemic dynamics on arbitrary metapopulation networks where the infection rate at each node depends on its population size. We have proved that the upper bound of endemic threshold  $\beta_c$ , previously obtained on the basis of an heterogeneous mean-field approximation, holds true even for arbitrary networks, regardless of this approximation. For an arbitrary network, the upper bound is achieved by assuming that a single node with the largest degree determines the onset of the endemic state. In addition, we have derived an improved upper bound for the endemic threshold, which decreases when the network possesses more links between large degree nodes. This structural property is directly related to the so-called rich-club phenomenon, observed in a variety of real-world networks. We have verified the resulting upper bounds with numerical simulations on network models, and on an empirical network, the US airports network, exhibiting the rich-club phenomenon and relevant to epidemic spreading.

Our finding that rich-club networks favor epidemic spreading is reminiscent of recent results

showing connections between endemic threshold and  $k$ -core structure for SIS dynamics on quenched contact networks (*i.e.*, networks between individuals) [43]. A  $k$ -core in a network is defined by the subgraph remaining after repeatedly removing nodes with degree smaller than  $k$  while recalculating nodes' degree on every node removal [31]. By definition, a  $k$ -core with a large  $k$  represents a highly-connected subgroup of nodes with large degrees. It was shown with numerical simulations [43] that nodes in the  $k$ -core sustain infection, a mechanism related to the finding that these nodes are influential initial spreaders for susceptible-infected-recovered (SIR) epidemic dynamics [1] on contact networks [44]. It should be noted that, prior to these physical studies, the impact of coreness in epidemic process was established in public health studies (*e.g.*, [45–47]). Although we considered metapopulation networks instead of contact networks, our finding suggests that a similar mechanism might occur in our case, as a highly-connected subgroup of nodes with large degrees sustains infection and favors onset of an endemic state.

This work opens interesting research questions related to epidemic dynamics on metapopulations networks. First, the upper bound  $\bar{\beta}$  is essentially obtained by estimating the principal eigenvector of the Jacobian matrix of the disease-free equilibrium, by using local network structure only (namely, node degree  $k_i$ ). As we have shown in the US airports network, this upper bound improves results obtained in the mean-field approach, but there is still room for improvement in order to predict the actual  $\beta_c$  measured in numerical simulations. Further improvement could be found by taking into account higher-order structural properties, such as community structure [48]. Second, in this paper, we have focused on the endemic threshold solely, but other quantities describe other properties of the dynamical process, often crucial from a theoretical point of view but also in practical public health. An important example is the fraction of infected particles at the stationary state, which requires nonlinear analytical approaches because the nonlinear terms associated with  $\rho_{S,i}\rho_{I,i}$  are not negligible and the linearization of the system fails in this regime.

## Acknowledgments

The authors thank to Naoki Masuda and Petter Holme for valuable discussions. This work was partly supported by Bilateral Joint Research Projects between JSPS, Japan, and F.R.S.–FNRS, Belgium. R.L. acknowledges support from FNRS, EU-FP7 project ‘Optimizr’, IAP DYSCO and ARC ‘Mining and Optimization of Big Data Models’. The data of the US airports network is downloaded from the Koblenz Network Collection, <http://konect.uni-koblenz.de/>.

## Appendix A: Degree assortativity coefficient $r$ for the random bimodal graph

From Eq. (51) in Ref. [49], we can rewrite the degree assortativity coefficient  $r$  as

$$r = 1 - \frac{\langle k \rangle \langle k^3 \rangle - \langle k \rangle \sum_k k^2 \bar{k}_{\text{nn}}(k) P(k)}{\langle k \rangle \langle k^3 \rangle - \langle k^2 \rangle^2}, \quad (\text{A1})$$

where  $\bar{k}_{\text{nn}}(k)$  is the average degree of the nearest neighbor nodes associated with a node with degree  $k$ ,

$$\bar{k}_{\text{nn}}(k) = \sum_{k'} k' P(k'|k). \quad (\text{A2})$$

We suppose that a node in the random bimodal graph has degree  $k_1$  and  $k_2$  with probability  $p(k_1) = a$  and  $p(k_2) = 1 - a$ , respectively ( $k_1 \geq k_2$ ). The degree correlation is governed by the conditional probability  $p(k_1|k_2)$  that is the probability with which a random link is connected to a node with  $k_1$  under the condition that the other end node has  $k_2$ . For the random bimodal graph, we derive:

$$\langle k^n \rangle = ak_1^n + (1-a)k_2^n \quad (n = 1, 2, 3, \dots), \quad (\text{A3})$$

$$\bar{k}_{\text{nn}}(k) = \begin{cases} k_1 \left\{ 1 - \frac{(1-a)k_2 P(k_1|k_2)}{ak_1} \right\} + k_2 \frac{(1-a)k_2 P(k_1|k_2)}{ak_1} & (k = k_1), \\ k_1 P(k_1|k_2) + k_2 (1 - P(k_1|k_2)) & (k = k_2). \end{cases} \quad (\text{A4})$$

When we substitute them into Eq. (A1) and simplify the equation, we obtain

$$r = 1 - \frac{(ak_1 + (1-a)k_2)P(k_1|k_2)}{ak_1}. \quad (\text{A5})$$

## Appendix B: Analytical expression of $\bar{\beta}$ and $\bar{\beta}_{\text{MFC}}$ for the random bimodal graph

For the random bimodal graph,  $\bar{\beta}$  (Eq. (32)) is equal to the smaller one in the two terms, i.e., either of the cases with  $k_c = k_1$  and  $k_c = k_2$ . We obtain  $E_c$  for  $k_c = k_1$  and  $k_c = k_2$  as

$$E_c = \begin{cases} MP(k_1, k_1) = M \frac{k_1 P(k_1|k_1)}{\langle k \rangle} P(k_1) = \frac{aNk_1}{2} \left\{ 1 - \frac{(1-a)k_2 P(k_1|k_2)}{ak_1} \right\} & (k_c = k_1), \\ M & (k_c = k_2), \end{cases} \quad (\text{B1})$$

where  $M$  is the total number of links and we use identity  $M = \langle k \rangle N/2$ . In addition,  $K_c$  is given by

$$K_c = \begin{cases} Nk_1 p(k_1) = aNk_1 & (k_c = k_1), \\ 2M & (k_c = k_2). \end{cases} \quad (\text{B2})$$

By combining Eqs. (B1) and (B2), we derive

$$1 - \frac{2E_c}{K_c} = \begin{cases} \frac{(1-a)k_2 P(k_1|k_2)}{ak_1} = (1-r) \frac{(1-a)k_2}{\langle k \rangle} & (k_c = k_1), \\ 0 & (k_c = k_2). \end{cases} \quad (\text{B3})$$

To derive the equation for the case with  $k_c = k_1$ , we use

$$P(k_1|k_2) = (1-r) \frac{ak_1}{\langle k \rangle}, \quad (\text{B4})$$

which is obtained by transforming Eq. (A5). Finally, we derive

$$\bar{\beta} = \min \left[ \left( \mu + (1-r) \frac{(1-a)k_2}{\langle k \rangle} D_I \right) \left( \frac{k_1}{\langle k \rangle} \rho \right)^{-\gamma}, \mu \left( \frac{k_2}{\langle k \rangle} \rho \right)^{-\gamma} \right]. \quad (\text{B5})$$

As a reference,  $\bar{\beta}_{\text{MFC}}$  (Eq. (34)) for the random bimodal graph is derived as

$$\bar{\beta}_{\text{MFC}} = \min \left[ \left( \mu + (1-r) \frac{(1-a)k_2}{\langle k \rangle} D_I \right) \left( \frac{k_1}{\langle k \rangle} \rho \right)^{-\gamma}, \left( \mu + (1-r) \frac{ak_1}{\langle k \rangle} D_I \right) \left( \frac{k_2}{\langle k \rangle} \rho \right)^{-\gamma} \right], \quad (\text{B6})$$

where we use

$$P(k_1|k_1) = 1 - P(k_2|k_1) = 1 - \frac{(1-a)k_2 P(k_1|k_2)}{ak_1}, \quad (\text{B7})$$

$$P(k_2|k_2) = 1 - P(k_1|k_2), \quad (\text{B8})$$

and Eq. (B4). In Fig. B1,  $\bar{\beta}$  and  $\bar{\beta}_{\text{MFC}}$  are plotted as a function of degree assortativity coefficient  $r$  while the parameters are set to the values we used in Sec. 3.2.3. For  $\gamma = 0.1$  and  $-1 \leq r \lesssim 0.5$ ,  $\bar{\beta} < \bar{\beta}_{\text{MFC}}$  holds. Otherwise, the two upper bounds are equivalent under these parameter settings.

## References

- [1] Anderson, R. M. and May, R. M. 1991 *Infectious diseases of humans: Dynamics and control*. Oxford University Press; Oxford.
- [2] Hanski, I. and Ovaskainen, O. 2000 The metapopulation capacity of a fragmented landscape. *Nature* **404**, 755–758. (doi:10.1038/35008063)
- [3] Colizza, V., Pastor-Satorras, R., and Vespignani, A. 2007 Reaction-diffusion processes and metapopulation models in heterogeneous networks. *Nat. Phys.* **3**, 276–282. (doi:10.1038/nphys560)
- [4] Zipf, G. K. 1949 *Human behavior and the principle of least effort: An introduction to human ecology*. Addison-Wesley Press; Oxford.
- [5] Gabaix, X. 1999 Zipf’s law for cities: an explanation. *Q. J. Econ.* **114**, 739–767. (doi:10.1162/003355399556133)
- [6] Carrothers, G. A. P. 1956 An historical bedew of the gravity and potential concepts of human interaction. *J. Am. Inst. Plann.* **22**, 94–102. (doi:10.1080/01944365608979229)
- [7] Saldaña, J. 2008 Continuous-time formulation of reaction-diffusion processes on heterogeneous metapopulations. *Phys. Rev. E* **78**, 012902. (doi:10.1103/PhysRevE.78.012902)
- [8] Colizza, V. and Vespignani, A. 2008 Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: theory and simulations. *J. Theor. Biol.* **251**, 450–67. (doi:10.1016/j.jtbi.2007.11.028)

- [9] Juher, D., Ripoll, J., and Saldaña J. 2009 Analysis and Monte Carlo simulations of a model for the spread of infectious diseases in heterogeneous metapopulations. *Phys. Rev. E* **80**, 041920. (doi:10.1103/PhysRevE.80.041920)
- [10] Saldaña J. 2010 Modelling the spread of infectious diseases in complex metapopulations. *Math. Model. Nat.* **5**, 22–37. (doi:10.1051/mmnp/20105602)
- [11] Juher, D. and Mañosa, V. 2014 Spectral properties of the connectivity matrix and the SIS-epidemic threshold for mid-size metapopulations. *Math. Model. Nat.* **9**, 108–120. (doi:10.1051/mmnp/20149207)
- [12] Smith, M. J., Telfer, S., Kallio, E. R., Burthe, S., Cook, A. R., Lambin, X., and Begon, M. 2009 Host-pathogen time series data in wildlife support a transmission function between density and frequency dependence. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 7905–7909. (doi:10.1073/pnas.0809145106)
- [13] Earn, D. J. 2000 A simple model for complex dynamical transitions in epidemics. *Science* **287**, 667–670. (doi:10.1126/science.287.5453.667)
- [14] Tildesley, M. J., Savill, N. J., Shaw, D. J., Deardon, R., Brooks, S. P., Woolhouse, M. E J., Grenfell, B. T., and Keeling, M. J. 2006 Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. *Nature* **440**, 83–86. (doi:10.1038/nature04324)
- [15] May, R. M. and Anderson, R. M. 1987 Transmission dynamics of HIV infection. *Nature* **326**, 137–142. (doi:10.1038/326137a0)
- [16] McCallum, H., Barlow, N., and Hone, J. 2001 How should pathogen transmission be modelled? *Trends Ecol. Evol.* **16**, 295–300. (doi:10.1016/S0169-5347(01)02144-9)
- [17] Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., and West, G. B. 2007 Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7301–7306. (doi:10.1073/pnas.0610172104)

- [18] Schläpfer, M., Bettencourt, L. M. A., Grauwin, S., Raschke, M., Claxton, R., Smoreda, Z., West, G. B., and Ratti, C. 2014 The scaling of human interactions with city size. *J. R. Soc. Interface* **11**, 20130789. (doi:10.1098/rsif.2013.0789)
- [19] Masuda, N. 2010 Effects of diffusion rates on epidemic spreads in metapopulation networks. *New J. Phys.* **12**, 093009. (doi:10.1088/1367-2630/12/9/093009)
- [20] Iggidr, A., Sallet, G., and Tsanou, B. 2012 Global stability analysis of a metapopulation SIS epidemic model. *Math. Popul. Stud.* **19**, 115–129. (doi:10.1080/08898480.2012.693844)
- [21] Lund, H., Lizana, L., and Simonsen, I. 2013 Effects of city-size heterogeneity on epidemic spreading in a metapopulation: a reaction-diffusion approach. *J. Stat. Phys.* **151**, 367–382. (doi:10.1007/s10955-013-0690-3)
- [22] Dorogovtsev, S. N., Goltsev, A. V., and Mendes, J. F. F. 2008 Critical phenomena in complex networks. *Rev. Mod. Phys.* **80**, 1275–1335. (doi:10.1103/RevModPhys.80.1275)
- [23] Anderson, R. M. and May, R. M. 1982 Directly transmitted infections diseases: control by vaccination. *Science* **215**, 1053–1060. (doi:10.1126/science.7063839)
- [24] Cohen, R., Erez, K., ben-Avraham, D., and Havlin, S. 2000 Resilience of the Internet to random breakdowns. *Phys. Rev. Lett.* **85**, 4626–4628. (doi:10.1103/PhysRevLett.85.4626)
- [25] Callaway, D. S., Newman, M. E. J., Strogatz, S. H., and Watts, D. J. 2000 Network robustness and fragility: percolation on random graphs. *Phys. Rev. Lett.* **85**, 5468–5471. (doi:10.1103/PhysRevLett.85.5468)
- [26] Pastor-Satorras, R. and Vespignani, A. 2001 Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203. (doi:10.1103/PhysRevLett.86.3200)
- [27] Pastor-Satorras, R. and Vespignani, A. 2002 Immunization of complex networks. *Phys. Rev. E* **65**, 036104. (doi:10.1103/PhysRevE.65.036104)



- [28] Chung, F. R. K. 1997 *Spectral Graph Theory*. American Mathematical Society; Providence, RI, USA.
- [29] Zhou, S. and Mondragón, R. J. 2004 The rich-club phenomenon in the Internet topology. *IEEE Comm. Lett.* **8**, 180–182. (doi:10.1109/LCOMM.2004.823426)
- [30] Colizza, V., Flammini, A., Serrano, M. A., and Vespignani, A. 2006 Detecting rich-club ordering in complex networks. *Nat. Phys.* **2**, 110–115. (doi:10.1038/nphys209)
- [31] Dorogovtsev, S. N., Goltsev, A. V., and Mendes, J. F. F. 2006 k-Core organization of complex networks. *Phys. Rev. Lett.* **96**, 040601. (doi:10.1103/PhysRevLett.96.040601)
- [32] de Oliveira, M. and Dickman, R. 2005 How to simulate the quasistationary state. *Phys. Rev. E* **71**, 016129. (doi:10.1103/PhysRevE.71.016129)
- [33] Ferreira, S. C., Ferreira, R. S., and Pastor-Satorras, R. 2011 Quasistationary analysis of the contact process on annealed scale-free networks. *Phys. Rev. E* **83**, 066113. (doi:10.1103/PhysRevE.83.066113)
- [34] Mata, A. S., Ferreira, S. C., and Pastor-Satorras, R. 2013 Effects of local population structure in a reaction-diffusion model of a contact process on metapopulation networks. *Phys. Rev. E* **88**, 042820. (doi:10.1103/PhysRevE.83.066113)
- [35] Barthélemy, M., Godrèche, C., and Luck, J.-M. 2010 Fluctuation effects in metapopulation models: percolation and pandemic threshold. *J. Theor. Biol.* **267**, 554–564. (doi:10.1016/j.jtbi.2010.09.015)
- [36] Molloy, M. and Reed, B. 1995 A critical point for random graphs with a given degree sequence. *Rand. Struct. Algorithms* **6**, 161–179. (doi:10.1002/rsa.3240060204)
- [37] Newman, M. E. J. 2010 *Networks: an Introduction*. Oxford University Press; Oxford.

- [38] Newman, M. E. J. 2002 Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701. (doi:10.1103/PhysRevLett.89.208701)
- [39] Newman, M. E. J. 2003 Mixing patterns in networks. *Phys. Rev. E* **67**, 026126. (doi:10.1103/PhysRevE.67.026126)
- [40] Hasegawa, T., Takaguchi, T., and Masuda, N. 2013 Observability transitions in correlated networks. *Phys. Rev. E* **88**, 042809. (doi:10.1103/PhysRevE.88.042809)
- [41] Opsahl, T. 2011 Why Anchorage is not (that) important: Binary ties and Sample selection. <http://toreopsahl.com/2011/08/12/why-anchorage-is-not-that-important-binary-ties-and-sample-selection/>
- [42] Kunegis, J. 2013 KONECT: the Koblenz Network Collection. In *Proceedings of the 22nd International Conference on World Wide Web Companion*, pp. 1343–1350, Rio de Janeiro, Brazil, ACM.
- [43] Castellano, C. and Pastor-Satorras, R. 2012 Competing activation mechanisms in epidemics on networks. *Sci. Rep.* **2**, 371. (doi:10.1038/srep00371)
- [44] Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., and Makse, H. A. 2010 Identification of influential spreaders in complex networks. *Nat. Phys.* **5**, 888–893. (doi:10.1038/nphys1746)
- [45] Laumann, E. O., Youm, Y. 1999 Racial/ethnic group differences in the prevalence of sexually transmitted diseases in the United States: A network explanation. *Sex. Transm. Dis.* **26**, 250–261. (doi:10.1097/00007435-199905000-00003)
- [46] Lowndes, C. M., Alary, M., Meda, H., Gnintoungbé, C. A. B., Mukenge-Tshibaka, L., Adjovi, C., Buvé, A., Morison, L., Laourou, M., Kanhonou, L., and Anagonou, S. 2002 Role of core and bridging groups in the transmission dynamics of HIV and STIs in Cotonou, Benin, West Africa. *Sex. Transm. Infect.* **78**, i69–i77. (doi:10.1136/sti.78.suppl\_1.i69)

- [47] Jolly, A. M., and Wylie J. L. 2002 Gonorrhoea and chlamydia core groups and sexual networks in Manitoba. *Sex. Transm. Infect.* **78**, i145–i151. (doi:10.1136/sti.78.suppl\_1.i145)
- [48] Fortunato, S. 2010 Community detection in graphs. *Phys. Rep.* **486**, 75–174. (doi:10.1016/j.physrep.2009.11.002)
- [49] Serrano M.A., Boguñá, M., Pastor-Satorras, R., and Vespignani, A. 2007 Correlations in Complex Networks. In: Caldarelli, G. and Vespignani, A. (eds.), *Large Scale Structure and Dynamics of Complex Networks: From Information Technology to Finance and Natural Science*, World Scientific, Singapore, pages 35–65 (chapter 3).

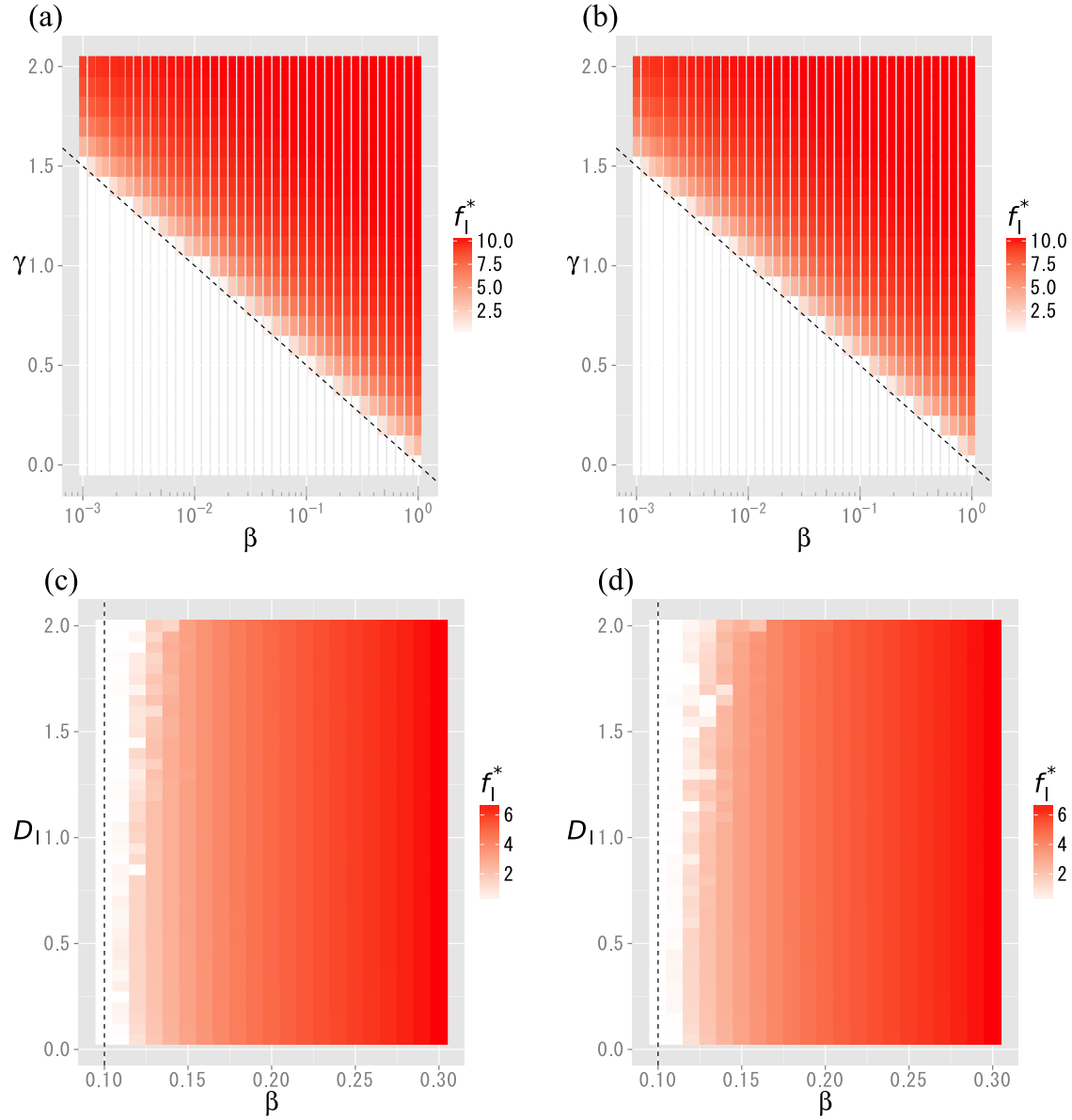


Figure 1: Results of numerical simulations on (a),(c) the random regular graph with  $k = 4$  and on (b),(d) the square lattice, both with  $N = 1024$ . Fraction of infected particles at the stationary state, denoted by  $f_I^*$ , are plotted as color maps on (a),(b)  $(\beta, \gamma)$ -plane and on (c),(d)  $(\beta, D_I)$ -plane. We set  $D_I = 1$  for (a),(c), and  $\gamma = 0.5$  for (b),(d). The dotted lines represent the theoretical prediction of  $\beta_c = \mu \rho^{-\gamma}$ .

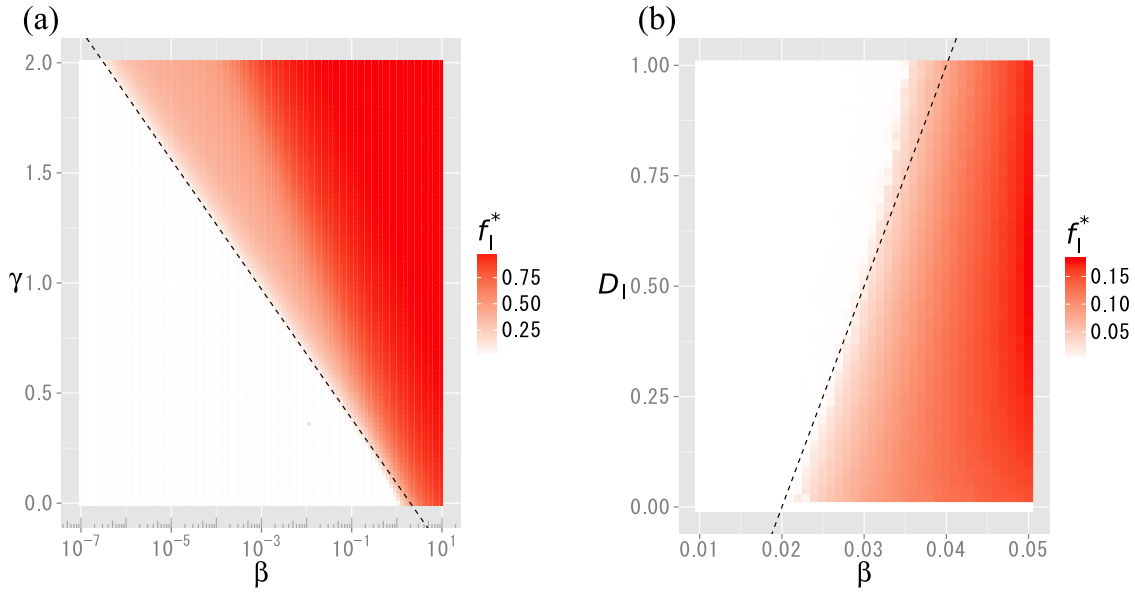


Figure 2: Results of numerical simulations on the wheel graph with  $N = 100$ . Fraction of infected particles at the stationary state, denoted by  $f_I^*$ , are plotted as color maps on (a)  $(\beta, \gamma)$ -plane and on (b)  $(\beta, D_I)$ -plane. We set  $D_I = 1$  for (a) and  $\gamma = 0.5$  for (b). Dotted lines represent the upper bound of  $\beta_c$  given by  $\bar{\beta} = (\mu + D_I)((N - 1)/\langle k \rangle \rho)^{-\gamma}$ .

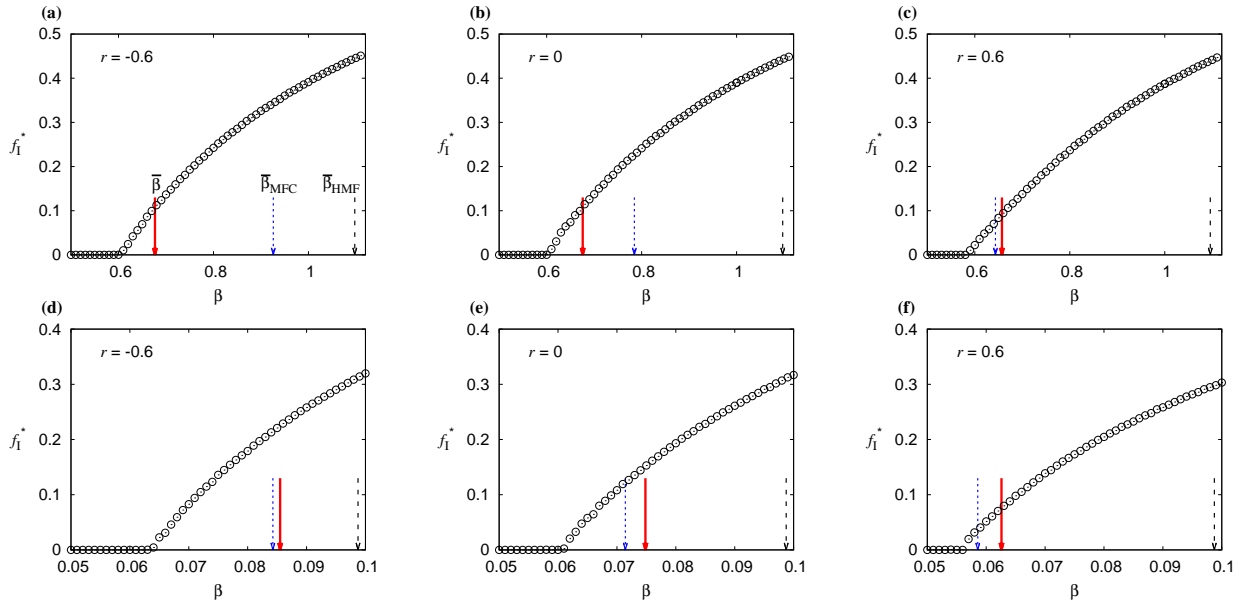


Figure 3: Results of numerical simulations on the random bimodal graph with degree correlation with  $(N, k_1, k_2, a) = (1000, 16, 2, 1/7)$ . Fraction of infected particle at the stationary state, denoted by  $f_I^*$ , are plotted as a function of  $\beta$  for (a),(b),(c)  $\gamma = 0.1$  and (d),(e),(f)  $\gamma = 0.5$ . We set the degree assortativity coefficient  $r$  to (a),(d)  $-0.6$ , (b),(e)  $0$ , and (c),(f)  $0.6$ . The vertical arrows indicate the three upper bounds: The solid ones are  $\bar{\beta}$ , the dotted ones  $\bar{\beta}_{\text{MFC}}$ , and the dashed ones  $\bar{\beta}_{\text{HMF}}$ .

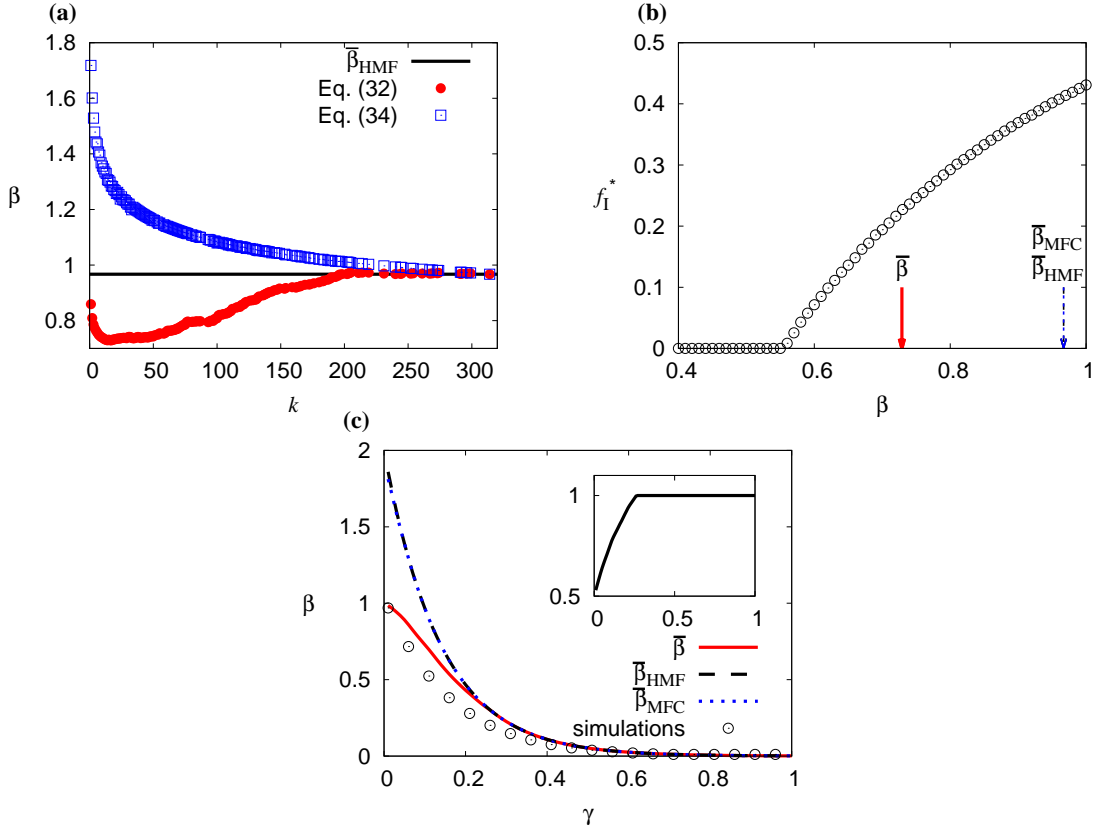


Figure 4: Results of numerical simulations on the US airports network. (a) Factor on the right hand side of Eq. (32) which is given by  $(\mu + (1 - 2E_c/K_c) D_I) (k_c \rho / \langle k \rangle)^{-\gamma}$  as a function of  $k_c$  (filled circles), and the factor on the right hand side of Eq. (34) which is given by  $(\mu + (1 - P(k_i|k_i)) D_I) (k_i \rho / \langle k \rangle)^{-\gamma}$  as a function of  $k_i$  (open squares). The solid horizontal line indicates  $\bar{\beta}_{\text{HMF}} = (\mu + D_I) (k_{\text{max}} \rho / \langle k \rangle)^{-\gamma}$ . (b) Fraction of infected particles at the stationary state, denoted by  $f_I^*$ , as a function of infection rate  $\beta$  (open circles). The solid, dashed, and dotted arrows point three upper bounds  $\bar{\beta}$ ,  $\bar{\beta}_{\text{HMF}}$ , and  $\bar{\beta}_{\text{MFC}}$ , respectively. We set  $D_I = 1$  and  $\gamma = 0.1$  for (a),(b). (c) Dependency of the endemic threshold on  $\gamma$ . The open circles indicate  $\beta$  above which  $f_I^* > 1/\rho$  holds true for the simulations. Inset: the ratio of  $\bar{\beta}$  to  $\bar{\beta}_{\text{HMF}}$  shown in the main panel.

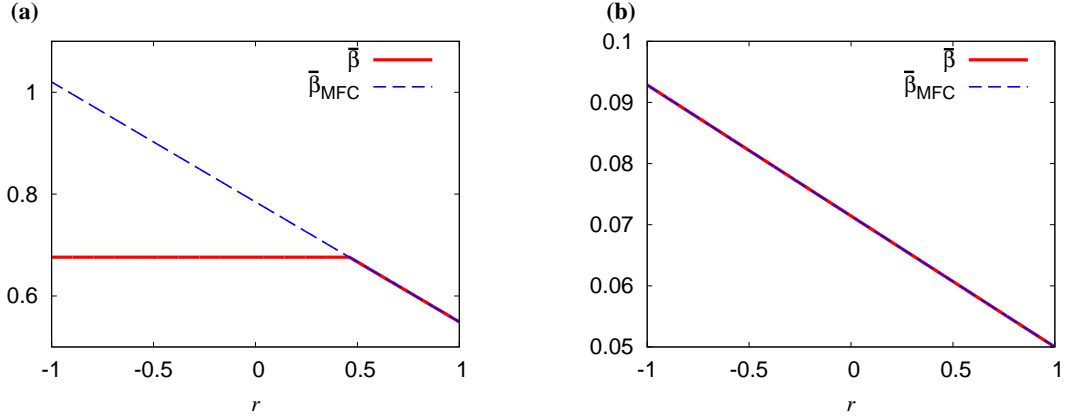


Figure B1: Upper bounds  $\bar{\beta}$  (Eq. (B5), solid) and  $\bar{\beta}_{\text{MFC}}$  (Eq. (B6), dashed) as a function of degree assortativity coefficient  $r$  for the random bimodal graph. We set (a)  $\gamma = 0.1$  and (b)  $\gamma = 0.5$ . Other parameters are fixed to  $(k_1, k_2, a, \rho, D_I) = (16, 2, 100, 1)$ , the same values as we used for the numerical simulations described in Sec. 3.2.3.